Determining Expert Profiles
(With an Application to Expert Finding)

Krisztian Balog and Maarten de Rijke
ISLA, University of Amsterdam

Motivation
- Searching an organization’s document repositories
- Finding the right person
  - expert on the topic
  - people you’d contact with questions on the topic
- From retrieving documents to retrieving objects

Related work
- Expert finding
  - “Who are the experts on topic X?”
  - Introduced at TREC in 2005
  - Given a query, return a ranked list of person names in response

- Problems:
  - Desired output should be more than a ranked list of person names
  - Context and evidence to help users

Expert’s profile
- Reverse of expert finding: “What does expert X know?”

- Topical profile
  - description of the areas in which she is an expert

- Social profile
  - description of her collaboration environment

The picture

query: authoring tools

Outline
- Introduction
- Topical profiles
  - Formal definition
  - Algorithms
- Evaluation
- An application to expert finding
- Conclusions
- Further work
Topical profile

Record of areas of skills and knowledge and the level of 'competency' in each.

\[ \text{profile}(ca) = \langle \text{score}(ca,ka_1), \ldots, \text{score}(ca,ka_n) \rangle \]

Knowledge Areas
- RDF Data Access
- No. Web Services Description

How to estimate \( \text{score}(ca,ka) \)?

Baseline

- Use an existing* expert finding method
  - \( p(\text{ca}|q) \) = probability of candidate \( ca \) being an expert given topic \( q \)
  - \( \text{rank}(ca,q) \) = position of \( ca \) on the ranked list of candidates given topic \( q \)
  - use knowledge area as the topic (\( q=ka \))

| Probability baseline | \( \text{score}(ca,ka) = p(\text{ca|ka}) \) |
|----------------------|------------------------------------------|
| Rank baseline        | \( \text{score}(ca,ka) = 1/\text{rank}(ca,ka) \) |


Method 1

- find documents that are relevant to the knowledge area
- sum up the relevance of those that are associated with the person

\[ \text{score}(ca,ka) = \sum_{d \in D_{ka}} \text{relevance}(d,ka) A(d,ca) \]

Method 2

- represent both the knowledge area and the candidate as a set of keywords
- ratio of co-occurring keywords is regarded as being the person’s competence

- each document \( d \) is represented as a set of keywords: \( KW(d) \)
  - keywords are extracted using TF-IDF

Method 2 cont’d.

- Represent knowledge area as a set of keywords:
  \( KW_{ka} = \bigcup_{d \in D_{ka}} KW(d) \)

- Represent individual as a set of keywords:
  \( KW_{ca} = \bigcup_{d \in D, A(d,ca)=1} KW(d) \)

- Estimate the person’s competence with the ratio of co-occurring keywords:

\[ \text{score}(ca,ka) = \frac{|KW_{ka} \cap KW_{ca}|}{|KW_{ka}|} \]

Filtering

- A knowledge area can be part of the candidate’s profile, if the person is among the top \( f \) ranked experts on that field
- Rank experts, using the profile scores
- Use these results to refine the output of the profiling method

\[ \text{score}'(ca,ka) = \begin{cases} 
\text{score}(ca,ka), & \text{if } |\{ ka' \mid \text{score}(ca',ka) < \text{score}(ca,ka) \}| < f \\
0, & \text{otherwise} 
\end{cases} \]
Evaluation
- Is “inverted expert finding” a viable solution?
- How do Method 1 and Method 2 perform?
- What is the impact of filtering?

Evaluation (2)
- TREC Enterprise 2005 platform
- W3C collection
  - Mixture of document types crawled from w3c.org (www, wikis, e-mail lists archive, etc.)
  - 330,000 documents, 5.7 GB
- List of 1092 candidate experts
  - unique ID, name, e-mail address(es)
  - 50 topics, and relevance judgments

Creating topics and relevance judgements
- Utilize: TREC 2005 topics are names of W3C working groups
  - Use working group names as knowledge areas
  - A knowledge area is part of a person’s profile, if the person is member of the corresponding working group

Results
- Both Method 1 and 2 outperform baseline
- Filtering: early precision enhancing effect

<table>
<thead>
<tr>
<th>Method</th>
<th>MAP</th>
<th>MRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (probability)</td>
<td>0.320</td>
<td>0.397</td>
</tr>
<tr>
<td>Baseline (rank)</td>
<td>0.203</td>
<td>0.244</td>
</tr>
<tr>
<td>Method 1</td>
<td>0.407</td>
<td>0.503</td>
</tr>
<tr>
<td>Method 2</td>
<td>0.397</td>
<td>0.486</td>
</tr>
<tr>
<td>Method 1 + filtering (f=15)</td>
<td>0.408</td>
<td>0.649</td>
</tr>
<tr>
<td>Method 2 + filtering (f=150)</td>
<td>0.383</td>
<td>0.511</td>
</tr>
</tbody>
</table>

Outline
- Introduction
- Topical profiles
  - Formal definition
  - Algorithms
- Evaluation
  - An application to expert finding
  - Conclusions
- Further work

An Application to Expert Finding
- Use profiles to improve on expert finding
- If a knowledge area ranks low on a person’s profile => push the candidate down on the list of experts
- Combine rankings

\[
\begin{align*}
\text{rank}_{kp}(ca, ka) &= \frac{1}{1 - \text{rank}_{kp}(ca, ka) \cdot \text{rank}_{kp}(ca, ka)} \\
A &= \frac{1}{\text{rank}_{kp}(ca, ka) \cdot \text{rank}_{kp}(ca, ka)} \\
B &= \lambda \cdot \text{rank}_{kp}(ca, ka) + (1 - \lambda) \cdot \text{rank}_{kp}(ca, ka)
\end{align*}
\]
Results

<table>
<thead>
<tr>
<th>#rel</th>
<th>MAP</th>
<th>MRR</th>
<th>P@5</th>
<th>P@10</th>
<th>P@20</th>
</tr>
</thead>
<tbody>
<tr>
<td>EF (baseline)</td>
<td>0.576 0.196</td>
<td>0.531 0.336</td>
<td>0.332 0.269</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method 1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A)</td>
<td>0.576 0.209*</td>
<td>0.659* 0.396*</td>
<td>0.326 0.267</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B) λ = 0.5</td>
<td>0.576 0.197</td>
<td>0.584* 0.376*</td>
<td>0.324 0.267</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method 2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A)</td>
<td>0.576 0.181</td>
<td>0.576* 0.340</td>
<td>0.292 0.242</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B) λ = 0.7</td>
<td>0.576 0.188</td>
<td>0.559* 0.344</td>
<td>0.306 0.254</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* effective in terms of early precision

Results (2)

<table>
<thead>
<tr>
<th>TREC 2005</th>
<th>MAP</th>
<th>MRR</th>
<th>P@5</th>
<th>P@10</th>
<th>P@20</th>
</tr>
</thead>
<tbody>
<tr>
<td>EF</td>
<td>0.196</td>
<td>0.531</td>
<td>0.336</td>
<td>0.332</td>
<td>0.269</td>
</tr>
<tr>
<td>EF+EP</td>
<td>0.209</td>
<td>0.659</td>
<td>0.396</td>
<td>0.326</td>
<td>0.267</td>
</tr>
<tr>
<td>TREC 2006</td>
<td>MAP</td>
<td>MRR</td>
<td>P@5</td>
<td>P@10</td>
<td>P@20</td>
</tr>
<tr>
<td>EF</td>
<td>0.328</td>
<td>0.506</td>
<td>0.395</td>
<td>0.408</td>
<td>0.377</td>
</tr>
<tr>
<td>EF+EP</td>
<td>0.466</td>
<td>0.851</td>
<td>0.661</td>
<td>0.587</td>
<td>0.495</td>
</tr>
</tbody>
</table>

* TREC’05 topics: effective in terms of early precision
* TREC’06 topics: very effective in all respects

Outline

- Introduction
- Topical profiles
  - Formal definition
  - Algorithms
  - Evaluation
- An application to expert finding
  - Conclusions
  - Further work

Conclusions

- Profiles provide context and evidence to users searching for expertise
- “Inverted” expert finding is not effective
- We proposed two methods and a filtering algorithm that significantly outperformed the baseline
- We applied profiling algorithms to enhance the performance of an expert finding method

Further work

- Further investigate the relation between expert finding and profiling
- More sophisticated models for creating profiles (Language Models)
- Evaluating and making use of social profiles
- How do these methods work on a different collection, e.g. university data set?

Questions

Krisztian Balog
kbalog@science.uva.nl
Social profile

Expert finding vs profiling

- \( p(\text{ca}|q) \) what is the probability of a candidate \( \text{ca} \) being an expert given the topic \( q \)?

- Expert profiling
  - \( p(\text{ka}|\text{ca}) \) what is the probability of a knowledge area \( \text{ka} \) being part of the candidate’s profile?
  - using knowledge area as a query (\( q=\text{ka} \)) \( \Rightarrow p(q|\text{ca}) \)

- Applying Bayes’ rule

\[
p(\text{ca}|q) = \frac{p(q|\text{ca})p(\text{ca})}{p(q)}
\]